Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics

Frank Raischel,[1] Teresa Scholz,[2] Vitor V. Lopes,[2] and Pedro G. Lind[1,*]

[1] *Center for Theoretical and Computational Physics, University of Lisbon, Av. Prof. Gama Pinto 2, 1649-003 Lisbon, Portugal*
[2] *Energy Systems Modeling and Optimization Unit, National Laboratory for Energy and Geology (LNEG), Estrada do Paço do Lumiar 22, 1649-038 Lisbon, Portugal*

(Dated: October 29, 2012)

Using a method for stochastic data analysis, borrowed from statistical physics, we analyze synthetic data from a Markov chain model that reproduces measurements of wind speed and power production in a wind park in Portugal. From the theoretical point of view we argue that our methods can be used to extract unknown functional relations between two variables. We first show that indeed our analysis retrieves the power performance curve, which yields the relationship between wind speed and power production and discuss how such procedure can be extended for extracting functional relationships between pairs of physical variables in general. Second, we show how specific features, such as the turbine rated wind speed or the descriptive wind speed statistics, can be related with the equations describing the evolution of power production and wind speed at single wind turbines.

PACS numbers: 02.50.Ga, 02.50.Ey, 92.70.Gt
Keywords: Energy systems, Environmental Research, Wind Turbines, Stochastic Systems

I. INTRODUCTION

The use of efficient and clean renewable energy sources is one of the major conditions required to achieve the important aim of sustainable development in modern societies[1]. Wind energy is one of such sources and wind turbines are being subject to intensive studies for improving their efficiency[2, 3]. Although the basic laws of atmospheric wind motion have been known for a long time, important problems such as turbulence, layering, and the statistics of extreme events remain poorly understood. A better understanding of these phenomena can help to construct energy convertion schemes that are both more efficient and robust. Here, robustness must be considered under two aspects: first, the occurrence of sudden changes in wind speed and direction can interrupt the process of energy conversion, meaning unreliability and a sudden slump in the electrical energy generated, which is seen as one of the major obstacles for the replacement of fossil and nuclear plants by wind energy sources. Second, these sudden changes introduce massive mechanical stresses which can lead to excessive wear or, ultimately, to the destruction of wind generators.

Wind flow is in general turbulent and non-homogeneous[4] with a non-negligible stochastic contribution. Therefore, in order to be able to construct more accurate models for its physical properties, one needs either accurate measurements of the wind speed on the length scales of wind turbines, or suitable models that can statistically reproduce these measured data. Since the wind turbines are driven by turbulent wind fields, the stochasticity of the wind fields transfers to stochastic dynamics of the wind turbine as a whole, of the loads on its structures and, last but not least, of the power output. Recently, a Markov chain model was used to reproduce wind measured data[5], based on the transition matrix and time propagators for the wind speed and direction together with the power production. Differently from previous first-order approaches[6], information from two and three step transition probabilities are considered.

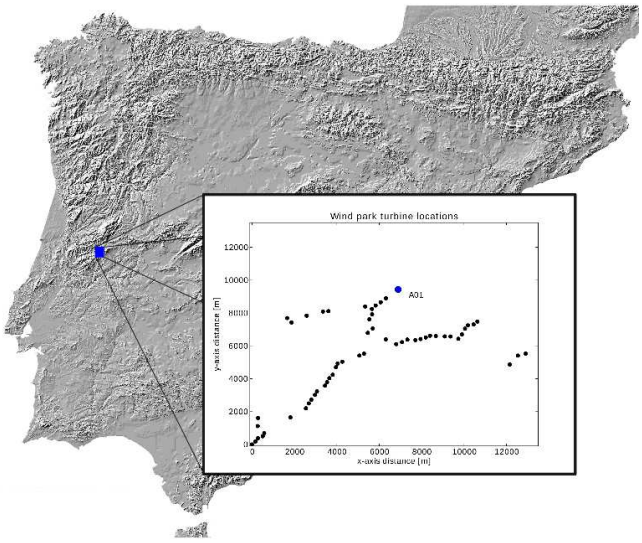

FIG. 1: Illustration of Iberia Peninsula indicating the position of the Portuguese wind park. In the inset one sees (bullets) the geographic location of each of the 57 wind turbines. In blue (AO1) one sees the wind turbine analyzed here.

In this paper, we aim at understanding the stochastic aspects of power production coupled to the wind velocity field. To that end, we use a methodology recently introduced by some of us[7], for uncovering optimal stochastic variables weakly[8] and strongly coupled[7], and adapt it with two purposes. First, to properly derive the functional relation of pairs of variables whose values are extracted from the Markov chain model for wind turbines. Second, to uncover specific features of the wind turbine and characterize the different working regions observed in power-speed plane. After this,

*Electronic address: plind@cii.fc.ul.pt

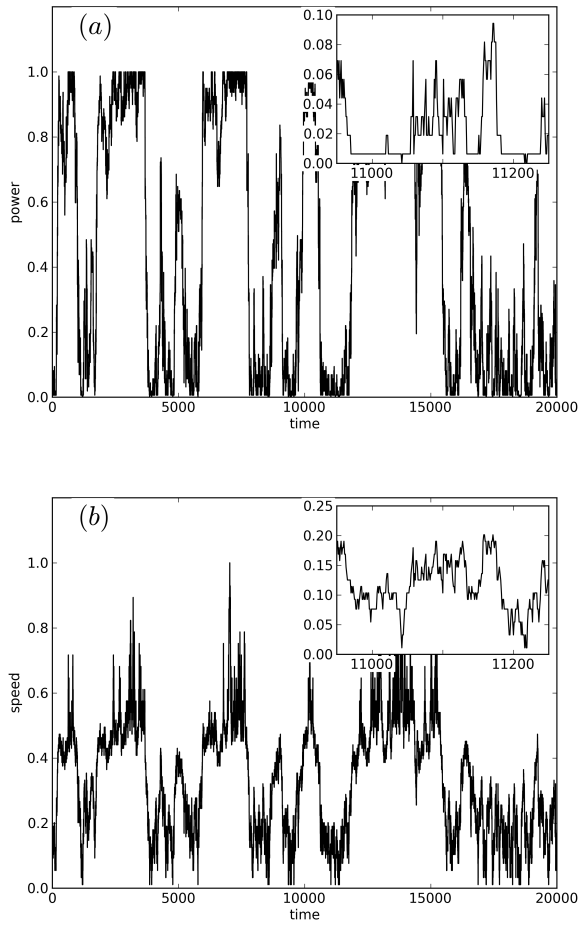

FIG. 2: Time series for **(a)** the power production P of the wind turbine, **(b)** the magnitude of the wind speed v. All data series were generated with the Markov chain model [5] described in Sec. II. Insets show each time-series in a short time period. All properties are normalized to the observed intervals $[0, P_{max}]$ and $[0, v_{max}]$ respectively.

we test our approach to uncover the functional dependence of the well-known performance curve, which describes the functional dependence of the power production with the wind speed. Whereas previous reports have pointed out the benefits of deriving the power curve from the drift field[9], we additionally take the diffusion field into account and find that this procedure creates additional insight.

We use the data sets generated by the Markov chain model described by Lopes et al. [5]. Using such synthetic data sets that properly reproduce the statistical features of empirical data sets, allows us to use data sets as large as needed for our analysis. Moreover, the Markov chain model behaves as a sort of filter for our modeling.

We start in Sec. II by describing the empirical data used to define the Markov chain model as well as the data generated with it. In Sec. III we describe our stochastic method for analyzing the data and in Sec. IV we apply it to analyze the performance curve of a wind turbine. In Sec. V we show that

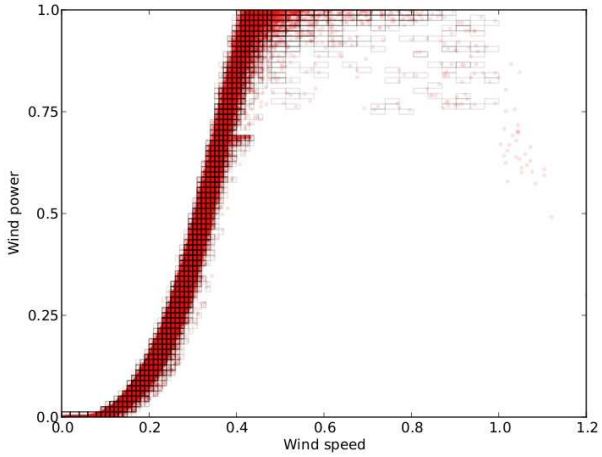

FIG. 3: Performance curve for one wind turbine in Pinhal Interior, Portugal. Circles show all the historical datapoints used in the Markov chain modeling [5] and the boxes display the state discretization.

the same method when applied separately to both wind speed and power production one is able to derive the performance curve. Further, the same analysis also provides insight concerning specific features of the turbine system studied. Section VI concludes this paper.

II. PROPERTIES AND GENERATION OF THE DATA SETS

The data analyzed in this manuscript was simulated from a set of measurements from a wind turbine in Portugal, region of Pinhal Interior. The measured properties were the power production P of the wind turbine and the wind speed v. The wind turbine was selected in an eolic park out of a total of 57 wind turbines. Figure 1 shows an overview of this Portuguese eolic park. The time increment between two successive measures was $\Delta t = 10$ minutes and the time period covered starts in January 1st 2009 and ends in December 31st 2010, yielding approx. 10^5 data points. It has to be remarked that this measurements are acquired directly from the top of wind turbine (nacelle) and might not be optimal for the reconstruction of the underlying physical processes for two reasons: first, the wind speed measurement is acquired at a point located downstream of the turbine blades and can neither account for the spatial extension and inhomogeneity of the wind field nor for its complex aerodynamical interaction with the turbine blades[10]. Secondly, the 10 minute sampling period of the historical dataset does not allows to resolve the time scales of either the turbulent interaction between wind and turbine, nor the quick action of the controller system response. Finally, missing data records, a low number of data points [11] or large sampling intervals [12], and periodicities due to the daily and seasonal variations in wind flow often hinders a direct stochastic analysis of these datasets. Thus, the challenge is to devise new methods that can make most use of the in-

formation present on these datasets—given that most of the data acquisition systems on existent wind farms are limited—with the aim of understanding the dynamic processes of the wind power generation, which hopefully can lead to economical benefits (production scheduling and maintenance).

To overcome some of these problems, we employ a reconstruction of the original data set through a Markov chain model. The Markov chain model is established using a joint discretization of the wind speed and power variables for the state definition and a multiple-step maximum likelihood estimator for the determination of the transition probabilities [5]. In our specific case within the range of each variable P and v we select 80 and 60 states (values). States without a realization in the time series are deleted. Figure 3 shows the discretization of the dataset where the combined states for power and speed are indicated with boxes and circles represent the historical dataset points projected into the speed-power plane.

Based on the Markov chain transition matrix $\mathbf{P}$ with $\mathbf{P}(i,j) = p_{i,j}$ being the probability of transition from state s_i to state s_j, the synthetic data sets were generated using the following Monte Carlo approach. We find the cumulative probability transition matrix $\mathbf{P}_{cum}$ with $\mathbf{P}_{cum}(i,j) = \sum_{k=1}^{j} p_{i,k}$ and select randomly an initial state s_i. A random number ϵ between zero and one is then uniformly selected and a new state s_{new} is chosen such that $P_{cum}(i,new) \geq \epsilon$. For details see Ref. [6]. Figure 2 shows the generated time series.

The resulting synthetic data series for power production and wind speed preserve their cumulative distributions and persistence statistics, namely the average duration of power production on a certain level (mean sojourn times). Power production and wind speed are presented as fractions of the maximum observed power P_{max} and wind speed v_{max} respectively, assuming therefore values between zero and one.

The synthetic datasets were generated with 2×10^6 data points and show stationary behavior, i.e. have constant moving averages (not shown).

This approach has several advantages over the direct analysis of historical datasets. First, high-quality data series of arbitrary length can be generated, which increases the accuracy of the Markov analysis. Secondly, the generated data are by construction Markovian, with the reconstruction through the Markov chain acting as a filter that removes both noise correlations and periodicities. Finally, non-Gaussian transition probabilities between the states are preserved, which enables to study them through higher Kramers-Moyal coefficients.

III. STOCHASTIC ANALYSIS OF WIND TURBINES

The co-evolution of two or more stochastic variables, such as wind speed and power production can be described through a system of coupled stochastic equations, each one defined by a deterministic contribution - the so-called drift - and stochastic fluctuations from possible stochastic sources. In this section we present the general framework to analyze our data and in the next section we apply it to the power production and wind speed variables.

For the general case of K stochastic variables, $X_1, \ldots, X_K$ the vector $\mathbf{X}(t) = (X_1(t), ..., X_K(t))$ defines the state of the system under study at each time instant t. The evolution of the state vector yields a stochastic trajectory in phase space and is given by the so-called Itô-Langevin equation[7, 13, 14]:

$$\frac{d\mathbf{X}}{dt} = \mathbf{h}(\mathbf{X}) + \mathbf{g}(\mathbf{X})\mathbf{\Gamma}(t), \tag{1}$$

where $\mathbf{\Gamma(t)} = (\Gamma_1(t), \ldots, \Gamma_K(t))$ is a set of K independent stochastic forces with Gaussian distribution fulfilling the following conditions: $\langle \Gamma_i(t) \rangle = 0$ and $\langle \Gamma_i(t)\Gamma_j(t') \rangle = 2\delta_{ij}\delta(t-t')$. Function $\mathbf{h} = \{h_i\}$ in Eq. (1) is the deterministic contribution, describing the physical forces which drive the system, while $\mathbf{g} = \{g_{ij}\}$ describes the amplitude of the stochastic sources of fluctuations $\mathbf{\Gamma}$[15].

The evolution of the stochastic variables in time yields a joint probability density function (PDF), $f(\mathbf{X})$, that evolves according to the so-called Fokker-Planck equation

$$\begin{aligned} \frac{\partial f(\mathbf{X},t)}{\partial t} &= -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left[D_i^{(1)}(\mathbf{X}) f(\mathbf{X},t)\right] \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j}\left[D_{ij}^{(2)}(\mathbf{X}) f(\mathbf{X},t)\right] \end{aligned} \tag{2}$$

where the functions $D_i^{(1)}$ and $D_{ij}^{(2)}$ are related with the functions h_i and g_{ij} above, namely

$$D_i^{(1)}(\mathbf{X}) = h_i(\mathbf{X}) \tag{3a}$$

$$D_{ij}^{(2)}(\mathbf{X}) = \sum_{k=1}^{N} g_{ik}(\mathbf{X}) g_{jk}(\mathbf{X}) \tag{3b}$$

and are usually called drift and diffusion functions respectively.

Drift and diffusion functions can be directly derived from observed or generated data [15, 16], and this fact is the heart of our framework. Indeed, the drift and diffusion functions of the underlying process are defined through conditional moments, namely [13]:

$$\mathbf{D}^{(k)}(\mathbf{X}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{\mathbf{M}^{(k)}(\mathbf{X}, \Delta t)}{k!} \quad , \tag{4}$$

where $\mathbf{M}^{(k)}$ are the first and second conditional moments ($k = 1, 2$). These conditional moments can be directly derived from the measured data as [15, 16] $M_i^{(1)}(\mathbf{X}, \Delta t) = \langle Y_i(t+\Delta t) - Y_i(t) | \mathbf{Y}(t) = \mathbf{X} \rangle$ and $M_{ij}^{(2)}(\mathbf{X}, \Delta t) = \langle (Y_i(t+\Delta t) - Y_i(t))(Y_j(t+\Delta t) - Y_j(t)) | \mathbf{Y}(t) = \mathbf{X} \rangle$ where $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_N(t))$ is the N-dimensional vector of measured variables and $\langle \cdot | \mathbf{Y}(t) = \mathbf{X} \rangle$ symbolizes a conditional averaging over the entire measurement period, where only measurements with $\mathbf{Y}(t) = \mathbf{X}$ are taken into account. Important conditions to hold are (i) the underlying process is stationary and (ii) the Markovian property is fulfilled.

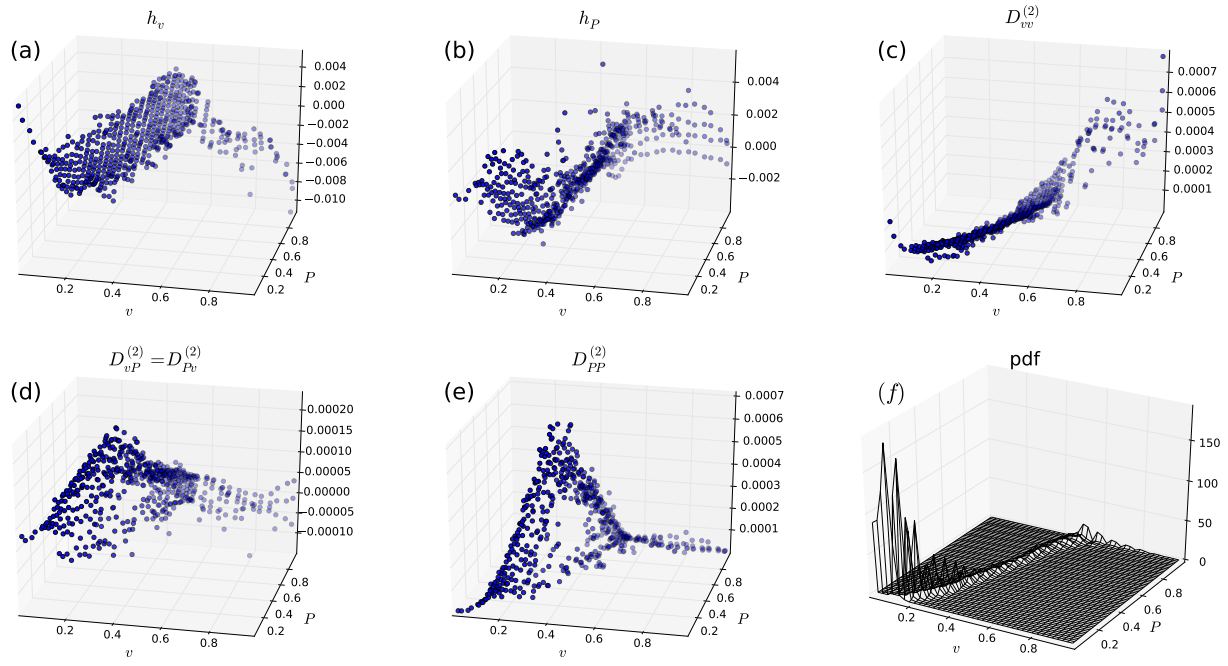

FIG. 4: The drift and diffusion coefficients defining the co-evolution of P and v: **(a)** h_v, **(b)** h_P, **(c)** $D^{(2)}_{vv}$, **(d)** $D^{(2)}_{vP} = D^{(2)}_{Pv}$ and **(e)** $D^{(2)}_{PP}$. The PDF for both variables is shown in **(f)**.

Numerically $\mathbf{h}$ and $\mathbf{G}$ are determined on a $n_1 \times ... \times n_N$ mesh of points in phase space, as a function of the variables X_i, using the drift and diffusion functions. Locally, at each mesh point, one can always diagonalize matrix $\mathbf{G}(\mathbf{X})$ and compute their K eigenvalues and K eigenvectors. As shown previously[7, 18, 19], this analysis provides information about the stochastic forces acting on the system. Namely, the eigenvalues indicate the amplitude of the stochastic force and the corresponding eigenvector indicates the direction toward which such force acts. In a previous work[7] we argued that to each eigenvector of the diffusion matrix one can associate one independent source of stochastic forcing Γ_i and thus the eigenvectors can be regarded as defining principal axes for stochastic dynamics. In particular, if one eigenvalue is very small compared to all the others, the corresponding stochastic force can be neglected. In the following sections we present a different implication of this principal stochastic component analysis, which emphasizes that the vanishing of one stochastic direction is in fact an indication of a strong functional dependence between the pair of variables being analyzed.

IV. THE PERFORMANCE CURVE FOR WIND TURBINES

In this section we focus on two variables solely, power production P and wind speed v. Since both series are stationary and Markovian we assume them to evolve according to the following equations:

$$\frac{dv}{dt} = h_v(v,P) + g_{vv}(v,P)\Gamma_1 + g_{vP}(v,P)\Gamma_2 \quad (5a)$$

$$\frac{dP}{dt} = h_P(v,P) + g_{Pv}(v,P)\Gamma_1 + g_{PP}(v,P)\Gamma_2 . \quad (5b)$$

In general, the six functions defining vector $\mathbf{h}$ and matrix $\mathbf{g}$ depend on both variables and describe the coupling between each other. Based on Eqs. (3), we derive both $\mathbf{h}$ and $\mathbf{g}$ from the functions $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, which, in turn, were extracted directly from the synthetic data-set by computing the corresponding conditional moments using Eq. (4). Note that, solving Eq. (3b) for computing the matrix $\mathbf{g}$ yields multiple solutions. If $\mathbf{g}$ is a solution then all matrices of the form $\tilde{\mathbf{g}} = \mathbf{gO}$ where $\mathbf{O}$ is an orthogonal matrix ($\mathbf{OO}^T = \mathbf{1}$) are also admissible solutions.

In our analysis we take $\mathbf{g}$ as the "square root" of matrix $\mathbf{D}^{(2)}$, i.e. we diagonalize $\mathbf{D}^{(2)}$ through a proper permutation matrix and since all eigenvalues are positive ($\mathbf{D}^{(2)}$ is positive definite) we take the square root of each eigenvalue and transform the matrix back.

Figures 4a-e show the five components of $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, i.e. the numeric results for both the drift and the diffusion coefficients computed directly from the generated P and v time-series. The large fluctuations in the region near to maximum power production and wind speed are due to lack of observations. Indeed, the joint PDF for P and v (Fig. 4f) shows that this region is poorly sampled.

To extract valuable information, next we treat these functions separately. Namely, we consider the drift vector field

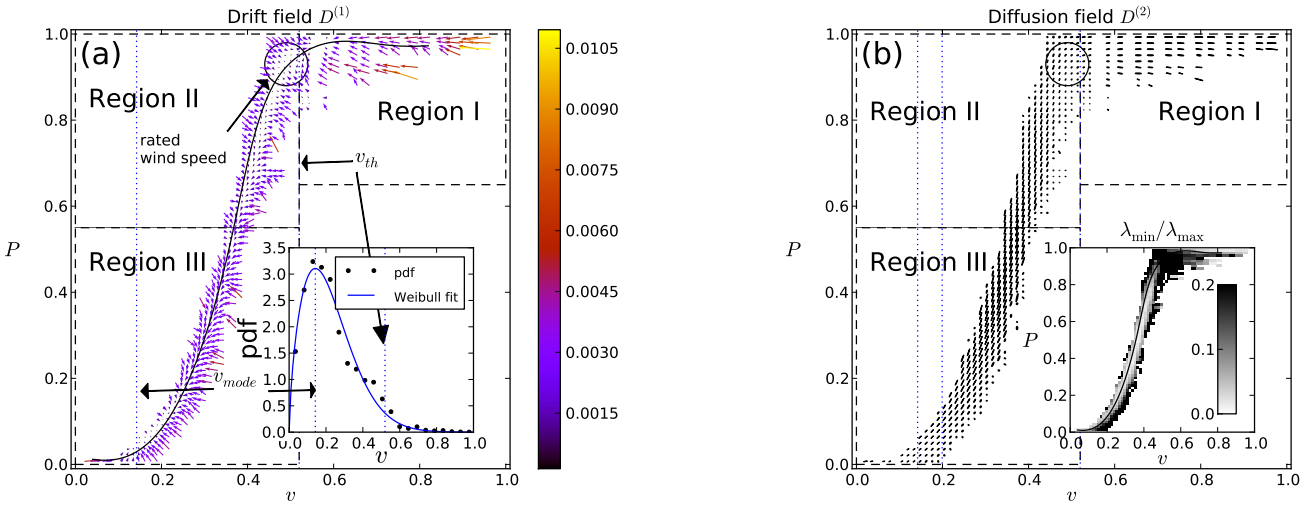

FIG. 5: **(a)** The drift vector $\mathbf{D}^{(1)}(P,v) = (h_P(P,v), h_v(P,v))$ (see Eq. (5)) in each (P,v) box used to generate the data (see Fig. 3). Three regions can be identified: Region I having slow dynamics and regions II and III with fast dynamics (see text). Interestingly, the fixed point spot in Region II coincides with the wind turbine rated speed (see text). The black curve indicates the performance curve and crosses the bins for which $D^{(1)}(v,P)$ vanishes. In the inset, the marginal probability density function of the wind speed v is well fitted by a Weibull distribution with scale parameter $\lambda \approx 0.25$ and shape parameter $k \approx 1.66$ and a mode $v_{mode} \approx 0.14 v_{max}$[17]. Velocities above $v_{th} = 0.52 v_{max}$ are rarely observed, which explain the observed drifts in Region I (see text). **(b)** Diffusion ellipses in the power production and wind speed state space, plotted together with the distribution of data pairs along the boxes shown in Fig. 3. At each box center, the corresponding diffusion ellipse is defined by the two orthogonal eigenvector of the diffusion matrix $\mathbf{D}^{(2)}$ computed at that center. The principal axis defining the ellipse are aligned along the eigenvectors with a length proportional to the corresponding eigenvalue. Along the performance curve which gives the functional dependence between both variables, P and v, the diffusion ellipses degenerate to a line segment tangent to the curve at each box center. This feature enables one to use the diffusion matrix of any set of variables for deriving their functional relationships (see text). In the inset one sees the ratio of both eigenvalues $\lambda_{min}/\lambda_{max}$, using a gray scale (0.2 for black, 0 for white).

(h_v, h_P) and the eigenvectors of the diffusion matrix associated to its eigenvalues λ_{max} and λ_{min}. Figure 5a shows the drift vector field in the power production and wind speed state space, restricted to the sampled region defined by the power production curve in Fig. 3. The solid black line is the performance curve computed from (P,v) the joint probability density function, shown in Fig. 4f, and defines the most likely power production for a given wind speed. Three different regions can be identified.

Region I is characterized by a large wind speed, i.e. above a "threshold" velocity v_{th} that exceeds the turbine rated wind speed. Here, the expected behavior of the wind turbine is to maintain the power production since there is an surplus of energy in the airflow. In this region, the performance curve is roughly constant at $\sim 0.95 P_{max}$. Still, positive power drifts are observed whenever the power production is below the performance curve. The wind speed drift is large in magnitude and always negative, i.e. the drift points towards lower wind velocities.

Region II is characterized by production levels above $0.5 P_{max}$ and wind speeds in $0.4 v_{max} \leq v \leq v_{th}$. Here, there is an broad area around the performance curve where the drift is relatively small in magnitude. An interesting feature is the location of the encircled area in Fig. 5a near the v_{th} where there is a change in the drift vector consistent with an attraction point (sink). This shows that the wind turbine rated speed was correctly selected for this location.

Region III is characterized by a power production below $0.5 P_{max}$ and for wind speed below $0.4 v_{max}$, where there is another attraction point at $v_{mode} \approx 0.14 v_{max}$. The separation between Region II and III coincides approximately with the boundary of the basins of attraction of each stable fixed point, i.e. the drift field in Region II and III tends to point away from the boundary towards the corresponding stable fixed point v_{mode}.

In previous works[4] the drift vector field around the performance curve is parallel to the power production axis. In Fig. 5a the vector fields tends to be inclined near the performance curve, because the data analyzed was sampled with a much smaller frequency, and therefore the time between successive measures is sufficiently large to observe the convergence to the stable fixed points.

Such observations can be more clearly understood by considering Fig. 5a together with the marginal PDF of the wind speed shown in its inset. The distribution of observed values for the wind speed typically follows a Weibull distribution, as is known from the literature[17].

The two dominant trends identified in region I are compatible with the expected behavior of the power production control system present on the wind turbine. For high values of wind speed, the controller action upon the blade aerodynamics is capable of sustaining the production level despite the

expected decrease of the wind speed. One has to consider, however, the time scales involved. With a 10min resolution of the original data, it is not possible to directly observe the rapid controller action on the blades, only the controller actions that occur on larger time scales, such as the rotation of the tower. However, even at large sampling times, the data set catches some events outside of the power curve and the subsequent conditional moments mirror the controller action that forces the system back on the curve.

A closer look in Region II enables one to identify a fixed point region ($\mathbf{D}^{(1)} \sim 0$) at high power production levels and wind speed $\mathbf{v} \sim \mathbf{0.5v_{max}}$. This speed value coincides with the rated wind speed, i.e. the speed for which the turbine was designed and at which it operates at an optimal regime. It can therefore be concluded from our analysis that the turbine has been well adjusted, and it remains to be seen if similar conclusions can be drawn when applying our method to arrays of turbines.

Another important application of our method deals with the diffusion matrix. As explained in the previous section, by diagonalizing the diffusion matrix at each point of the phase space one is able to determine the two eigendirections for diffusion. Being orthogonal to each other, these two eigendirections define an ellipse with major and minor axis proportional to the corresponding eigenvalue. Figure 5b shows the diffusion ellipses in phase space. Region I is characterized by the largest ellipses indicating very large fluctuations, while there is an area in region II, which present small fluctuations and corresponds to the fixed point areas identified in the drift. In the high-slope region of the power curve, the ellipses degenerate, i.e. one eigenvalue is negligible when compared to the other ($\lambda_{min}/\lambda_{max} \sim 0$). The inset of Fig. 5b shows in a gray scale the quotient $\lambda_{min}/\lambda_{max}$ between the smallest and the largest eigenvalue. White corresponds to zero quotient, while values in $[0.2 \ldots 1]$ are colored in black. Clearly, a white curve can be identified in the inset which follows the performance curve shown in Fig. 3.

V. DERIVING THE PERFORMANCE CURVE FROM UNIVARIATE STOCHASTIC DYNAMICS

The inset of Fig. 5b shows that along the performance curve one eigenvalue is typically much larger than the other. As we will argue next, such feature is an indication that in fact P is a function of v, which in the case of power production and wind speed yields the performance curve drawn in Fig. 3 and 5.

Interestingly, we will see that analyzing both wind speed and power production separately enables one to extract valuable insight about the full dynamics and behavior of the wind turbine with the atmospheric wind. In fact, the 2D analysis of the performance curve summarized in Fig. 5 can indeed be accessed through a one dimensional stochastic analysis of each variable P and v separately.

To see this one first takes P and v as two general variables fulfilling Eqs. (5) and observes that if $P \equiv P(v)$ there are not two independent stochastic forces, but only one, yielding for

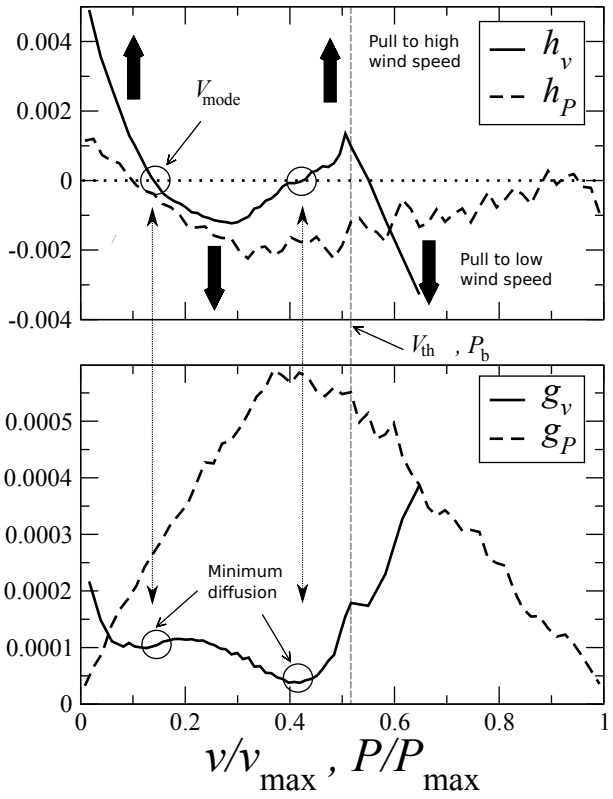

FIG. 6: Uncovering properties of wind turbines by analyzing data series of wind speed v and power P separately (see text and Eqs (6) and (7)). Horizontal axis indicates the value of v/v_{max} for h_v and g_v and the value of P/P_{max} for h_P and g_P.

v[14]

$$\frac{dv}{dt} = \tilde{h}_v(v) + \tilde{g}_v(v)\Gamma, \tag{6}$$

and for P

$$\frac{dP}{dt} = \tilde{h}_P(P) + \tilde{g}_P(P)\Gamma, \tag{7}$$

where functions $\tilde{h}$ and $\tilde{g}$ are of course different from the drift and diffusion functions defined above in Eqs. (5), since only one variable is taken into consideration for the stochastic motion equation.

Figure 6 shows the drift and diffusion of both the wind speed and of power production determined for the model described by equations Eqs. (6,7). The drift of the wind speed, h_v has three zeros. These zeros correspond to three fixed points, two stable ($v \simeq v_{mode} \simeq 0.14v_{max}$ and $v \simeq v_{th}$), and one unstable at $v \simeq 0.4v_{max}$. Thus, for wind speed below $\simeq 0.4v_{max}$, the airflow is unstable and unsuited for power production, while wind speeds above $\simeq 0.4v_{max}$ promote power production. The first zero of h_v indicates approximately the mode of the wind speed value distribution - compare with inset in Fig. 5a - and the other two zeros mark the transition between two different regions identified above in Fig. 5. The transition between Region I and Region II is

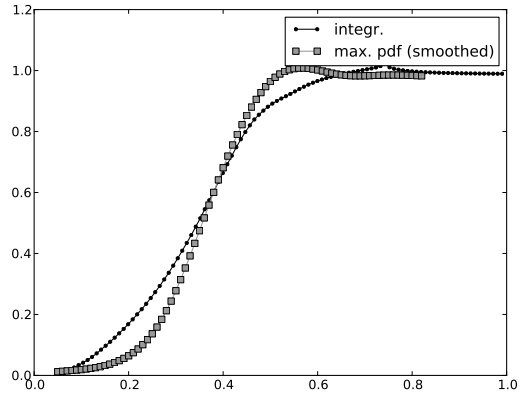

FIG. 7: Obtention of power production P as function of wind speed v by integration (see Eq. (8)). Both Eqs. (9) are fulfilled, analyzing both series P and v separately (see Eqs. (6) and (7)).

marked by v_{th}. The transition between Region II and III is more subtle and deals with the zero at $0.4v_{max}$ and with the functions for both variables: it is located at the value P_b which corresponds to $v \simeq 0.4v_{max}$, cf. Fig. 5a. Moreover, Fig. 6a also shows that positive drifts are located for small wind speed (up to v_{mode}) and for Region II (see Fig. 5), while Fig. 6b shows the transition value P_b of power production is close to the maximum of g_P.

The transition between Region III and Region II is located at a minimum of the diffusion g_v for the wind velocity. As shown in Fig. 6b, here the drift changes to a positive value. In other words, above $v \simeq 0.4v_{max}$ the expected change of the wind speed is towards higher values. The drift reaches a maximum on Region II prior to a steep change towards negative values. In this region, the wind speed values are not in the range of the extreme weather conditions and are also not as frequent as the lower wind speed values. However, it is frequent enough to be associated with a commonly repeated pattern, i.e. the daily pattern seasonality, in which the airflow is mainly induced by thermal differences. This pattern is responsible for most of the power production in this wind turbine and the main reason for a second attraction point at v_{th}. For higher wind speed values, the drift changes again to negative values (Region I) since very high wind speed is usually of short duration, i.e. extreme wind gusts.

Bringing all the above observations into account, one concludes that there is a strong agreement between the regions defined in the context of Fig. 5 and the sign of the wind speed drift.

Having analyzed separately both properties, v and P, we end this section by showing that from the drift and diffusion coefficients, h_v, g_v, h_P and g_P, one indeed obtains the functional dependence between power production and wind speed. To that end we assume that Eq. (6) holds for v and that the other variable P is an exclusive function $P(v)$ of v. Thus, we can take the Ito-Taylor expansion[14] of its differential

$$\begin{aligned} dP(v) &= P(v+dv) - P(v) \\ &= \frac{dP}{dv}dv + \tfrac{1}{2}\frac{d^2P}{dv^2}dv^2 + \mathcal{O}(dv^3) \\ &= \left(\frac{dP}{dv}\tilde{h}_v + \tfrac{1}{2}\frac{d^2P}{dv^2}\tilde{g}_v^2\right)dt + \frac{dP}{dv}\tilde{g}_v dw \end{aligned} \quad (8)$$

using the differential $dv = \tilde{h}_v(v)dt + \tilde{g}_v(v)dw$. Therefore, identifying

$$\tilde{h}_P = \frac{dP}{dv}\tilde{h}_v + \tfrac{1}{2}\frac{d^2P}{dv^2}\tilde{g}_v^2 \quad (9a)$$

$$\tilde{g}_P = \frac{dP}{dv}\tilde{g}_v \quad (9b)$$

which can be solved with respect to the two derivatives of $P(v)$ yielding the numerical integration scheme as follows:

$$\begin{aligned} P(v+\Delta t) &= P(v) + \frac{dP}{dv}|_{v,P(v)}\Delta v + \frac{1}{2}\frac{d^2P}{dv^2}|_{v,P(v)}(\Delta v)^2 + \mathcal{O}((\Delta v)^3) \\ &= P(v) + \frac{\tilde{g}_P(P(v))}{\tilde{g}_v(v)}\Delta v + \frac{1}{2}\frac{\tilde{h}_P(P(v))\tilde{g}_v(v) - \tilde{h}_v(v)\tilde{g}_P(P(v))}{(\tilde{g}_v(v)^3}(\Delta v)^2 + \mathcal{O}((\Delta v)^3). \end{aligned} \quad (10)$$

Figure 7 shows the integration of $dP(v)$ for the condition $P_0(v_0) = 0$ for $v_0 = 0$. The deviations can be attributed to the fact that Eqs. (6) and (7) are strictly only valid in those regions where the eigenvalues of the diffusion matrix show a large difference between them, $\lambda_{min} \ll \lambda_{max}1$, cf. inset of Fig 5 where $0.2 \leq v \leq v_{th}$. In addition, it also holds only on the performance curve, and applying Eqs. (6) and (7) therefore also neglects the asymmetry of the drift functions with respect to this curve.

VI. DISCUSSION AND CONCLUSIONS

Investigating a wind turbine from a commercial wind park, we report the reconstruction of the stochastic performance curve in the variables wind speed and power production, using both drift and diffusion coefficients. These coefficients, describing the respective deterministic and stochastic interactions of wind field, turbine aerodynamics, and controller action, are estimated from a synthetic time series reconstructed from a Markov Chain model of the original measurement data. We argue that this reconstruction is highly superior to a direct

evaluation of the measurements.

As a main finding, we present the fact that the reconstruction of the power curve using both drift and diffusion coefficients uncovers additional information not visible in an analysis of the drift field alone[9], even though we are using measured data of a very low measurement rate as model input. Specifically, our analysis reveals the existence of various distinct regions in the wind speed–power production plane

In addition, we have been able to reconstruct the power curve from the drift and diffusion coefficients, using a method which should be able to uncover functional relationships between stochastic variables in a wide range of experimental setups.

We our convinced that our approach is helpful in obtaining a better understanding of the complex dynamics that determines power production in wind turbines, and that it can be useful in optimizing the output. It is planned to extend this methodology in order to consider coupled systems of wind turbines in nearby locations.

Acknowledgments

The authors thank Matthias Wächter for useful discussions and GENERG, SA. FR (SFRH/BPD/65427/2009), TS (SFRH/BD/86934/2012) and PGL (*Ciência 2007*) thank Fundação para a Ciência e a Tecnologia for financial support, also with the support under PEst-OE/FIS/UI0618/2011 and FCOMP-01-0124-FEDER-016080.

[1] W.M. Adams, "The Future of Sustainability: Re-thinking Environment and Development in the Twenty-first Century." Report of the IUCN Renowned Thinkers Meeting, 2931 January 2006.

[2] 46. U.S. Department of Energy; Energy Efficiency and Renewable Energy "20% Wind Energy by 2030", available at `http://www.windpoweringamerica.gov`.

[3] A. Rauh and J. Peinke, "A Phenomenological Model for the Dynamic Response of Wind Turbines to Turbulent Wind", Journal of wind engineering and industrial aerodynamics, Bd. 92, 159-184 (2004).

[4] J. Gottschall, J. Peinke, Journal of Physics: Conference Series **75** 012045 (2007).

[5] V.V. Lopes, T. Scholz, A. Estanqueiro, and A.Q. Novais, "On the use of Markov chain models for the analysis of wind power time-series", Environment and Electrical Engineering (EEEIC), 2012 11th International Conference on, pages 770-775, 2012.

[6] A.D. Sahin and Z. Sen, "First-order Markov chain approach to wind speed modeling", J. Wind Eng. Ind. Aerodyn. **89**, 263-269 (2001).

[7] V. V. Vasconcelos, F. Raischel, M. Haase, J. Peinke, M. Wächter, P.G. Lind and D. Kleinhans, Phys. Rev. E **84** 031103 (2011).

[8] F. Raischel, A. Russo, M. Haase, D. Kleinhans and P.G. Lind, Physics Letters A **376**, 2081-2089 (2012).

[9] E. Anahua, S. Barth, J. Peinke, Wind Energy **11**, 219 (2008).

[10] M. Wächter, H. Heielmann, M. Hölling, A. Morales, P. Milan, T. Mücke, J. Peinke, N. Reinke, and P. Rinn, Journal of Turbulence **13**, N26 (2012).

[11] David Kleinhans, Phys. Rev. E **85**, 026705 (2012).

[12] S.J. Lade, Phys. Rev. E **80** 031137 (2009).

[13] H. Risken, *The Fokker-Planck Equation*, (Springer, Heidelberg, 1984).

[14] C. W. Gardiner, *Handbook of stochastic Methods*, (Springer, Germany, 1997).

[15] R. Friedrich, J. Peinke, M. Sahimi, and M.R.R. Tabar, Phys. Rep. **506** 87 (2011).

[16] P.G. Lind, M. Haase, F. Boettcher, J. Peinke, D. Kleinhans and R. Friedrich, Phys. Rev. E **81** 041125 (2010).

[17] J.A. Carta and P. Ramírez and S. Velázquez, " A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands", Renewable and Sustainable Energy Reviews**13**(5), 933-955 (2009).

[18] J. Gradišek, R. Friedrich, E. Govekar and I. Grabec, "Examples of Analysis of Stochastic Processes Based on Time Series Data", Meccanica, **38**, 33 (2003).

[19] A. M. van Mourik, A. Daffertshofer, and P. J. Beek, Biological cybernetics **94**, 233 (2006).